

The University of Bradford Institutional Repository

<http://bradscholars.brad.ac.uk>

This work is made available online in accordance with publisher policies. Please refer to the repository record for this item and our Policy Document available from the repository home page for further information.

To see the final version of this work please visit the publisher's website. Access to the published online version may require a subscription.

Link to publisher's version: <http://dx.doi.org/10.1098/rsta.2013.0384>

Citation: Booth J, Vazquez S, Martinez-Nunez E et al (2014) Recent applications of boxed molecular dynamics: a simple multiscale technique for atomistic simulations. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 372(2021)

Copyright statement: © 2014 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/3.0/>, which permits unrestricted use, provided the original author and source are credited.



Recent applications of Boxed Molecular Dynamics: a simple multiscale technique for atomistic simulations

Jonathan Booth¹, Saulo Vazquez², Emilio Martinez-Nunez², Alison Marks³, Jeff Rodgers,⁴
David R. Glowacki^{4,5*}, and Dmitrii V. Shalashilin^{1*}.

¹School of Chemistry, University of Leeds, Leeds LS2 9JT, UK; UK, ²Departamento de Química Física and Centro Singular de Investigación en Química Biológica y Materiales Moleculares, Campus Vida, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain, ³School of Life Sciences, University of Bradford, Bradford BD7 1DP, UK, ⁴School of Chemistry, University of Bristol, Bristol BS8 1TS, ⁵Department of Chemistry, Stanford University, Stanford, CA, 94305

*Corresponding authors: *drglowacki@gmail.com*, *d.shalashilin@leeds.ac.uk*

Abstract

In this article we briefly review the Boxed Molecular Dynamics (BXD) method, which allows analysis of thermodynamics and kinetics in complicated molecular systems. BXD is a multiscale technique, in which thermodynamics and long-time dynamics are recovered from a set of short-time simulations. In this article, we review previous applications of BXD to peptide cyclization, diamond etching, solution-phase organic reaction dynamics, and desorption of ions from self-assembled monolayers (SAMs). We also report preliminary results of simulations of diamond etching mechanisms and protein unfolding in AFM experiments. The latter demonstrate a correlation between the protein's structural motifs and its potential of mean force (PMF). Simulations of these processes by standard molecular dynamics (MD) is typically not possible, since the experimental timescales are very long. However, BXD yields well-converged and physically meaningful results. Compared to other methods of accelerated MD, our BXD approach is very simple; it is easy to implement, and it provides an integrated approach for simultaneously obtaining both thermodynamics and kinetics. It also provides a strategy for obtaining statistically meaningful dynamical results in regions of configuration space that standard MD approaches would visit only very rarely.

1. Introduction

Boxed Molecular Dynamics¹⁻³ (BXD) is a simple and straightforward technique which extends the time scale of atomistic Molecular Dynamics (MD) simulations and facilitates simulation of rare events. In BXD’s simplest implementation, we assume that a chemical reaction or some other atomistic physical process can be described by some reduced description of the system’s $3N$ dimensional configuration space – e.g., a reaction coordinate or an appropriate order parameter. It is then possible to split this coordinate into several boxes and subsequently lock the dynamics in each box by inverting the velocity of the trajectory in the direction of the reaction coordinate (or order parameter) every time the trajectory hits a boundary between two subsequent boxes. Then one can calculate the rate constant of exchange between the boxes simply by calculating average time between the “hits”.

$$\begin{aligned} k_{m,m+1} &= \frac{h_{m,m+1}}{t_m} \\ k_{m,m-1} &= \frac{h_{m,m-1}}{t_m} \end{aligned} \tag{1}$$

where $h_{m,m-1}$ and $h_{m,m+1}$ are the number of “hits” of the left and right boundary of the m -th box and t_m is the time spent in this box. After accumulating sufficient statistics the trajectory is allowed into the neighbouring box where the procedure is repeated. Eventually a set of rate constants for exchange between the boxes is accumulated and the dynamics is reduced to a set of kinetic equations.

$$\begin{aligned}
\frac{dn_1(t)}{dt} &= -(k_{12} + k_{10})n_1(t) + k_{21}n_2(t) \\
\frac{dn_2(t)}{dt} &= k_{12}n_1(t) + k_{32}n_3(t) - (k_{21} + k_{23})n_2(t) \\
&\dots \\
\frac{dn_m(t)}{dt} &= k_{m-1,m}n_{m-1}(t) - k_{m,m-1}n_m(t)
\end{aligned} \tag{2}$$

or in the matrix form

$$\frac{d\mathbf{n}(t)}{dt} = \mathbf{M}\mathbf{n}(t) \tag{3}$$

Then the MD simulation of the system is replaced by the solution of the Master Equation (2)

$$\mathbf{n}(t) = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}\mathbf{n}(0) \tag{4}$$

where $\mathbf{n}(0)$ contains the initial conditions for each box, \mathbf{U} is the eigenvector matrix obtained from diagonalization of \mathbf{M} , where $\mathbf{\Lambda}$ is a diagonal matrix whose elements, $\Lambda_{jj} = e^{\lambda_j t}$, are determined by λ , the eigenvalue vector corresponding to \mathbf{M} .

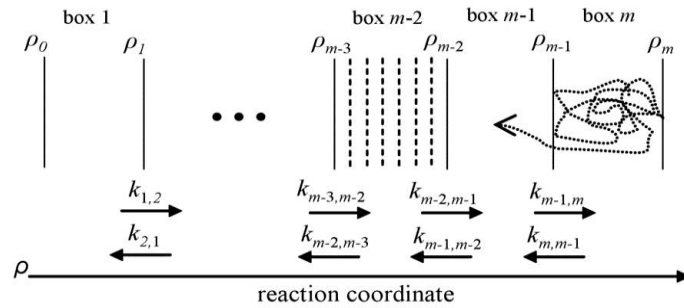


Figure 1: Schematic of how BXD works with a reaction coordinate, ρ , split into m boxes

BXD relies on the assumption that the motion within the box is stochastic and that sequential “hits” and velocity inversions are uncorrelated – i.e., the time between the “hits” must be bigger than the correlation time. In general this is not the case. A trajectory reflected from a boundary can sometimes turn back rather quickly. One needs to use a simple technique described in Ref² in order to remove these short time correlated events. The requirement of uncorrelated dynamics also imposes a restriction on the box size – i.e., it should be bigger than the so-called correlation length, which is the length at which a trajectory loses the memory of its initial conditions. For large anharmonically coupled systems, the correlation time and correlation length are usually quite short. The procedure developed in Ref² is based on analysis of the distribution of first passage times. It removes any contribution to the rate coefficients from velocity inversions separated by very short time and improves the quality of the kinetic results.

With the box-to-box rate coefficients in hand, it is possible to estimate the free energy profile along the reaction coordinate:

$$K_{n-1,n} = \frac{k_{n-1,n}}{k_{n,n-1}} = \exp\left(-\frac{\Delta G_{n-1,n}}{k_B T}\right) \quad (5)$$

A quantity which is closely related to the free energy is the Potential of Mean Force (PMF). Using BXD, it has been shown¹ that the free energy (or PMF) profile can actually be estimated at a resolution much higher than the box size. The procedure involves splitting boxes into smaller regions and calculating the time spent within those smaller regions (enclosed by dashed lines in Fig.1). This gives a high resolution free energy profile within each box. These profiles can then be stitched together using an exact procedure which yields the global free energy profile along the entire reaction coordinate. We have also developed a

post-analysis procedure for calculating kinetic rate coefficients between small boxes in cases where that information is useful.³

For us, BXD has its origin in the Intramolecular Dynamics Diffusion Theory (IDDT)⁴⁻⁸ which has demonstrated that long time reaction rates can be recovered from a set of short time MD simulations⁸. Longer timescale dynamics may be calculated from the set of rate constants obtained from short-time simulations within each box. This connection between short and long timescales is what makes BXD a multiscale method. In IDDT, the dynamics was not restricted by the boundaries of the box. The idea of the box was introduced in Ref⁹, where the first BXD simulation of bond fission was carried out using only two boxes with a single boundary between them. Locking the trajectory within a box enables the system to visit the regions of high free energy and low probability, where an unrestricted trajectory would rarely visit.

Rare event acceleration is an important part of modern MD simulations. Over the last few years, a number of rare event acceleration strategies have been proposed, many of which share similarities with BXD. Milestoning¹⁰⁻¹⁵ and particularly Milestoning with Voronoi tessellations (MVT),^{13, 16} are the most closely related to the BXD technique and it is worthwhile discussing connections between BXD and milestoning. Both use short time dynamics to recover local kinetic information about the motion along the reaction coordinate in order to solve a kinetic state-to-state master equation. The original milestoning procedure required accurate sampling of the equilibrium distribution at the milestone surfaces located along the reaction coordinate. This is very similar to the earlier IDDT procedure,^{7, 8} with the primary difference in the post-processing step: milestoning analysis relies on the milestone-to-milestone passage time, and IDDT analysis relies on a diffusional coefficient. In MVT^{13, 16}, a modified milestoning procedure has been proposed wherein the dynamics initiated

between a set of milestones are then locked using a collision rule at the boundary so that they cannot escape, as in Ref (9). Perhaps the most important difference between the velocity inversion scheme used by BXD and the collision rule of MVT is that the BXD scheme conserves energy, linear momentum, and angular momentum. So long as the distance between a BXD inversion boundary and transition state is larger than the characteristic system decorrelation length, then the accelerated dynamics obtained using BXD give meaningful statistics that allow detailed analysis of non-equilibrium dynamical properties in the post-transition state region of a given molecular system's phase space. (see ref 2) Indeed, these features of BXD have proven critical to many of the studies which are outlined in this paper, where BXD was used to investigate non-equilibrium energy relaxation in NVE trajectories¹⁷⁻¹⁹

BXD and MVT also differ in the post-processing step – i.e., how they interpret the MD simulation results obtained from within the boxes. BXD uses the trajectory information to calculate box-to-box transition state theory rate constants and MVT uses the information to calculate transition times for moving from border to border. As has been shown in Ref.² calculating either of the two is not entirely straightforward and requires thinking about dynamical decorrelation. In BXD the decay trace needs to be corrected by removing fast events where trajectories return to the boundary. This gives decay traces which are better approximated by a single exponent and box-to-box dynamics better described in terms of a single rate constant. MVT, on the other hand, gives flat nonexponential decays at short times. When these are removed, then the BXD and MVT decay traces are identical. The foundation for MVT is based on ideas from transition path sampling, and it is exact in the limit that the correct set of milestones are utilized.¹⁴ BXD has its foundations in standard transition state theory. It is exact in the limit that the system is ergodic, and a correct set of dividing surfaces

have been utilized. Based on BXD's origins in transition state theory, we have shown how to simply and exactly renormalize the BXD statistics to obtain thermodynamics¹ and kinetics³ to arbitrary histogram resolution, which is not part of milestoning or MVT.

BXD also shares similarities with a number of other rare event acceleration methods. For example, BXD in its simplest form⁹ is related to Hyperdynamics,^{20, 21} which similarly introduces constraints into a molecular system's configuration space to encourage it to visit regions of low probability. BXD also has similarities to Umbrella sampling²²⁻²⁵ with the primary difference that the former uses rectangular boxes instead of parabolic energy restraints. Compared to these methods, BXD's main advantage is that it is the only technique which is capable of simultaneously extracting high-resolution thermodynamic and kinetic information. BXD's formulation is very straightforward, based upon a simple rewriting of the classical transition state theory formula, which is common in chemistry. In common with all of the aforementioned techniques, BXD does not address the question of finding a good reaction coordinate or order parameter; its usage relies on being able to define these in a sensible way. In many instances, definition of an appropriate reaction coordinate or order parameter is a substantial challenge, but there are a range of systems where these quantities are reasonably well defined. If a good reaction coordinate exists, then BXD provides long timescale dynamics/kinetics and PMF profiles in regions of very high energies.

In this article we review several recent and ongoing applications of the BXD technique. First we review the simulation of peptide loop formation dynamics where interesting and unusual power law kinetics observed in experiments²⁶⁻²⁸ have been reproduced on timescales spanning several orders of magnitude – i.e., from picoseconds to microseconds. Second, we report new results from our simulations of ion diffusion through a monolayer on a metal surface, where residence times of hours were recovered in agreement with experiment²⁹⁻³¹.

Third, we give a brief summary of how BXD has been used to provide insight into the elementary chemical mechanisms of diamond etching. Fourth, we describe how BXD has been used to accelerate rare event sampling in studies of solution phase bimolecular reactions. Finally, we report preliminary BXD results of the free energies obtained from pulling apart several small proteins, all of which were previously investigated experimentally using Atomic Force Microscopy (AFM)³²⁻³⁴. For these systems, we observe that the slope of the preliminary PMFs calculated as a function of peptide extension (distance between the peptide's C and N termini) appear to be in reasonable agreement with the experimentally observed AFM forces required to unfold the proteins. For proteins composed of alpha-helices, the force required along the extension coordinate is substantially lower than for those composed of beta-sheets, in qualitative agreement with experimental observations. We also discuss future developments and applications.

2. Applications of BXD

2.1 Peptide loop formation

Volk and Hochstrasser previously carried out a series of detailed studies investigating the kinetics of polypeptide cyclization. Using a laser pulse, their experiments photolytically cut a solvated polypeptide loop (in which the two termini of the polypeptide backbone were connected via a disulfide bond), generating a thiyl radical pair localized on each end of the peptide.²⁶⁻²⁸ The radical concentration was monitored by time resolved spectroscopy spanning picoseconds to milliseconds^{27, 28}. The measured absorbance of the peptide occurs predominantly from the thiyl radicals, decreasing when the radical pairs recombine to form

the bond between the two ends of the polypeptide chain, i.e. when the peptide loop is reformed. In these experiments, the instantaneous rate constant²⁶⁻²⁸

$$k_{inst} = -\frac{1}{n_{open}(t)} \frac{dn_{open}(t)}{dt} \quad (6)$$

was determined to have the form

$$k \approx A t^{-\alpha} \quad (7)$$

where k has units of ns^{-1} , $A = 3.2$ and $\alpha = -0.93 \pm 0.02$, with similar values observed for loop formation dynamics in a range of peptides. Remarkably, this time dependence, which is very close to simple $k=1/t$ dependence, holds for nine orders of magnitude in time, from picoseconds to milliseconds, and was essentially the same for a short polypeptide²⁷ and a 100-residue protein²⁸. To rationalise these experimental results, we performed the first fully atomistic simulations of the system, using BXD to investigate several model peptides with initial conditions corresponding to those of experiment, i.e. at $t=0$ the two ends of the peptide are in close contact. In our simulations, the reaction coordinate was taken to be the peptide extension. This coordinate was split into anywhere between 15-20 boxes, with sizes ranging from 1 – 2 Å. Box-to-box rate constants were obtained using a decorrelation algorithm which we recently described.² The procedure developed in the Ref³ was used to extract ‘high resolution’ rate coefficients from the low resolution box-to-box rate coefficients. Having obtained the high-resolution rate coefficients, solution (4) of the kinetic master equation (3) was obtained. Figure 2 shows BXD results obtained for one of the peptides investigated in this study, namely 10-ALA.

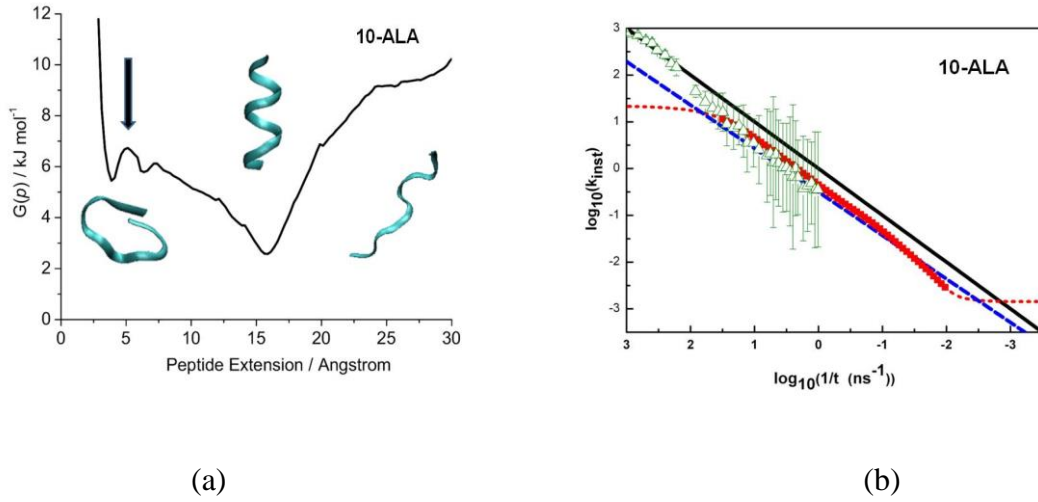


Figure 2: BXD results obtained from 10-ALA. Panel (a) shows the PMF as a function of extension coordinate. The arrow indicates the transition state for going from the extended conformation to the closed loop form. Panel (b) shows the results of BXD simulation (thick red line) together with the outcome of unbiased short-time MD simulations (triangles). The two plateaus shown by the red dashed line correspond to the smallest and largest eigenvalues of the matrix \mathbf{M} in Eq (3) and therefore represent the limits between which the BXD master equation results are valid. The result is also compared with a $1/t^\alpha$ power law, with $A = 1$, $\alpha = 1$ (black line) and $A = 3.2$, $\alpha = 0.93$ (blue dashed line).

Frame (a) depicts the PMF as a function of the reaction coordinate. Frame (b) presents the instantaneous rate constant of cyclization of the 10-ALA peptide. Triangles show the rate constant obtained by unbiased MD simulation. Using a serial implementation of the CHARMM program suite³⁵ on a local computing cluster, these unbiased simulations only permit reaction events on a picosecond timescale. The red line demonstrates the outcome of the BXD simulation. The plateaus shown by the red dotted line indicate the limitations of BXD – i.e., the BXD results are only reliable for times t where $t_{max} > t > t_{min}$ where t_{max} and t_{min} are the longest and shortest timescales obtained from the eigenvalue spectrum, respectively. As can be seen from the Fig 2, BXD reproduces a power law very similar to those obtained from fitting of the experimental data, which are shown by blue and black lines. Using BXD, atomistic simulations can be used to derive kinetic information on the microsecond timescale, far beyond timescales which are accessible by unbiased MD.

Similar results were obtained for a number of model peptides and it was shown that for all of them the instantaneous rate follows the power law (7) for the rate very accurately. The behaviour appears to be largely independent of the structure of peptides and it can be explained on the basis of the solution (4) of the master equation (3), giving the following form for the time dependence of the extended peptide:

$$n_{open}(t) = 1 - n_0(t) = \sum_i F(\lambda_i) e^{\lambda_i t} = \sum_i F(k_i) e^{-k_i t} \quad (8)$$

where $n_0(t)$ is the population of the box 0, which corresponds to loop closure, and $F(\lambda_i)$ is a pre-exponential factor which weights the contribution of a particular eigenvalue to the overall time-dependent population of extended peptide, $n_{open}(t)$. The rate constants $k_i = -\lambda_i$ of the exponents in (4) are related to the eigenvalues of the kinetic matrix \mathbf{M} in (3). It was noticed in the ref³ that the coefficients $F(\lambda_i)$ depend strongly on the initial populations of the boxes $n_i(0)$. Our analysis focussed on the eigenvalue spectrum weighted by the corresponding coefficients $F(\lambda_i)$. The shape of this weighted spectrum determines whether the phenomenological kinetics follow a power law or a standard form of a single exponent. In the limit that a single value in the weighted spectrum dominates, typical single exponential kinetics are observed. In the limit that there are no dominant values in the weighted spectrum, power law kinetics arise. In the immediate aftermath of experimental bond-breaking, all population is concentrated in the “boxes” adjacent to the loop formation transition state. This results in a scenario where the weighted spectrum has no dominant values, and all the rate coefficients k_i have characteristic times $t_i = 1/k_i$, that are equally probable, giving rise to the time dependence (7) for the instantaneous rate (6).

These results can be understood with the help of a simple non-scientific analogy. Suppose that a large group of hikers randomly choose their routes through a landscape similar to that

shown in figure 3, which provides a schematic representation of the energy landscape of a peptide. The randomness of each hikers' path means that there will be a distribution of available paths, each of which has a unique characteristic time for returning home. At time t_2 , for example, those hikers who took a route with characteristic time t_1 have already returned, while those on a longer route t_3 are still far away. Thus, at a given time t only the routes with the rate constant close to $k=1/t$ contribute, which explains the origin of the $1/t$ law in (8). An important feature of the Scheme depicted in figure 3, which distinguishes the “returning hiker” analogy from numerous existing random walk models, is that the all routes are initialized and terminated in roughly the same place. This means that shorter routes contribute similarly to longer routes.



Figure 3. Routes selected randomly on a landscape. At time t_2 the hikers who took the route t_3 are still walking but the hikers on the short route t_1 have already returned. Therefore at the moment t_2 the instantaneous rate constant of return is equal to $1/t_2$. If the start and end points of all routes are in close proximity, then there is a large number of short routes, which is not the case when start and end point are distant from each other.

The example of peptide cyclization not only demonstrates that BXD can produce reliable thermodynamic and kinetic information on very long time scales, (allowing one to reproduce

experiments), but that it also helps to provide simple insight to the qualitative mechanisms that give rise to the experimental observations.

2.2 Desorption of ions from the monolayer

As has been shown in the previous example, BXD is capable of reproducing the dynamics of peptides on timescales of microseconds where the barriers along the PMF are relatively small. An important additional question is whether BXD can accurately treat processes with much larger energy barriers along the PMF, where the dynamical timescales are substantially slower. In a recent study³⁶ the dynamics of the desorption of silyl ions $(\text{CH}_3)_2\text{SiNCS}^+$ and SiNCS^+ , trapped in the self-assembled monolayer of perfluorinated alkanethiol on gold (F-SAM) has been investigated. The monolayer with $(\text{CH}_3)_2\text{SiNCS}^+$ adsorbed on it is shown in Figure 4a. In the experiment²⁹⁻³¹ a monolayer was bombarded with each of these ions (separately) during about one hour forming ions adsorbed just beneath the surface of the monolayer (see Fig.4(a)). Subsequently, the F-SAMs were analyzed by high energy collisions of Xe^+ ions, which induce the ejection of surface species (sputtering). The sputtered ions were detected by mass spectrometry. It was shown that the larger ion $(\text{CH}_3)_2\text{SiNCS}^+$ was observed in the mass spectrum after several days, in stark contrast to the complete lack of the smaller SiNCS^+ ion. The interpretation given in the experimental studies is that the methyl groups increase the attractive and steric interactions between the ion and the F-SAM, which facilitate the entrapment of the ion inside the monolayer and decrease its desorption rate.

To address whether atomistic MD simulations were able to provide any insight into these experimental observations, we carried out BXD simulation of the ions diffusing inside of the

monolayer towards the surface with subsequent desorption. For a reaction coordinate we used the distance from the ion center of mass to the surface. The residence time was estimated using the kinetic master equation. The BXD module in the CHARMM code was used and BXD box-to-box rate constants were obtained. The rate constants were corrected by the previously discussed decorrelation procedure, which excludes the correlated velocity inversions separated by very short time. The rate constants included an additional correction, designed to account for the presence of image charge at the gold surface. This procedure allows one to compute the survival probability for ions diffusing from the monolayer to the vacuum, which is shown at the Figure 4a, yielding the residence time of 2 hours for $(\text{CH}_3)_2\text{SiNCS}^+$ and 3 seconds for the more compact SiNCS^+ ion, which explains why SiNCS^+ is not seen in the mass spectrum at long times.

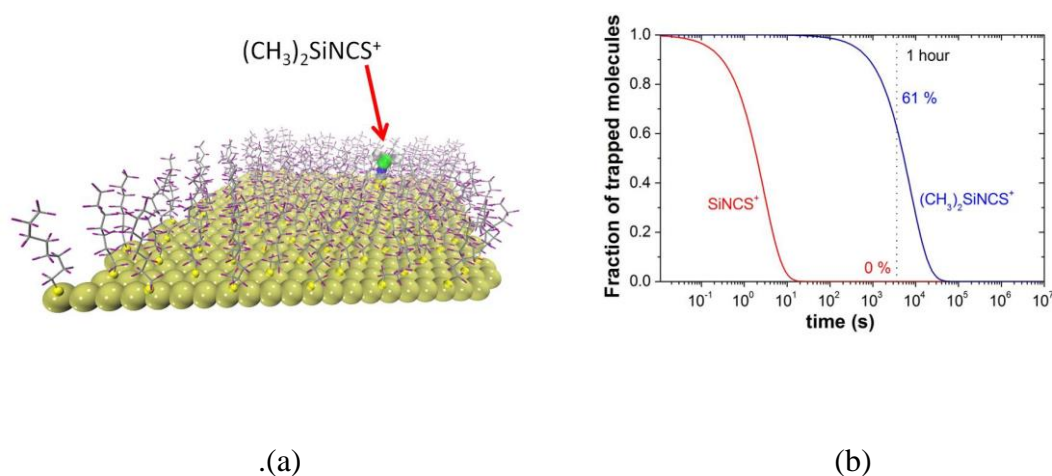


Figure 4 (a) Monolayer of perfluorinated alkanethiol on gold with adsorbed ion; (b) survival probability of SiNCS^+ (red) and $(\text{CH}_3)_2\text{SiNCS}^+$ (blue) in the monolayer calculated as a fraction of the trapped molecules.

This system provides a nice demonstration of the fact that BXD is capable of reaching timescales which are many orders of magnitude longer than the timescales typically recoverable by atomistic simulations. This is because of BXD's multiscale approach:

atomistic simulations yield box-to-box rate constants, which are used as subsequent input for kinetic analysis, covering much longer times.

2.3 Heterogeneous Chemistry at the surface of Diamond

BXD is also being applied to better understand the elementary chemical mechanisms that guide diamond etching – a process wherein solid-state diamond is converted to its gas-phase atomic constituents³⁷. In particular, BXD is being used to understand the kinetics and thermodynamics for dissociation of methyl radicals from diamond surfaces, according to the scheme in Figure 5a-b. Spontaneous thermal dissociation of C-C bonds at the surface of diamond has been suggested as a possible etching mechanism. Preliminary PMFs along the C-CH₃ stretching coordinate using BXD in conjunction with a molecular dynamics simulation setup shown in Figure 5c (at 1000K) suggest free energy barriers to methyl dissociation which are on the order of 38 – 46 kcal mol⁻¹, depending on temperature. The magnitude of these free energy barriers results is consistent with etching of diamond surfaces occurring as a result of spontaneous loss of methyl radicals.

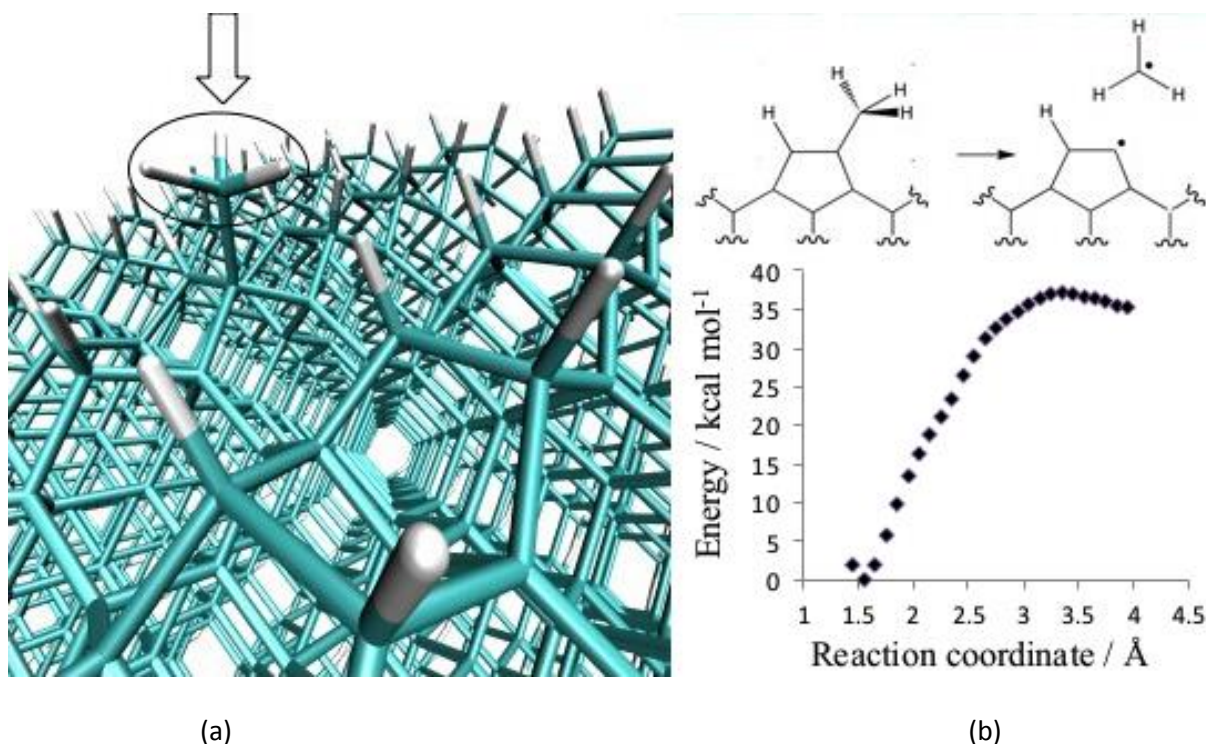


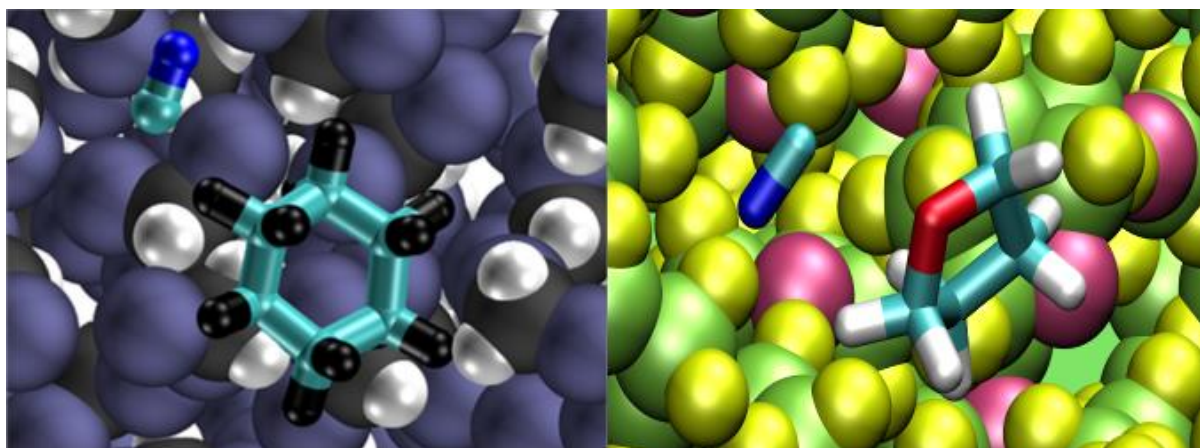
Figure 5: Panel (a) shows the atomistic MD setup upon which we carried out BXD sampling to determine PMFs for dissociation of the surface -CH_3 group, which is highlighted by the arrow; panel (b) shows a schema for a simple C-C bond dissociation etching mechanism, creating a nascent methyl group; panel (c) shows the 1000K PMF for methyl dissociation from a diamond surface, calculated using BXD

2.4 Solution Phase Dynamics Studies

Another application where BXD has proven invaluable is in the area of solution phase reaction dynamics³⁸. In particular, this method has permitted atomic level insight into solution-phase bimolecular reaction dynamics for the following hydrogen abstraction systems: (1) $\text{CN} + \text{cyclohexane}$ [$\text{CN} + \text{C}_6\text{H}_{12} \rightarrow \text{HCN} + \text{C}_6\text{H}_{11}$] in dichloromethane solvent^{17, 18, 39} (2) $\text{CN} + \text{tetrahydrofuran}$ [$\text{CN} + \text{C}_4\text{H}_8\text{O} \rightarrow \text{HCN} + \text{C}_4\text{H}_7\text{O}$] in tetrahydrofuran solvent⁴⁰

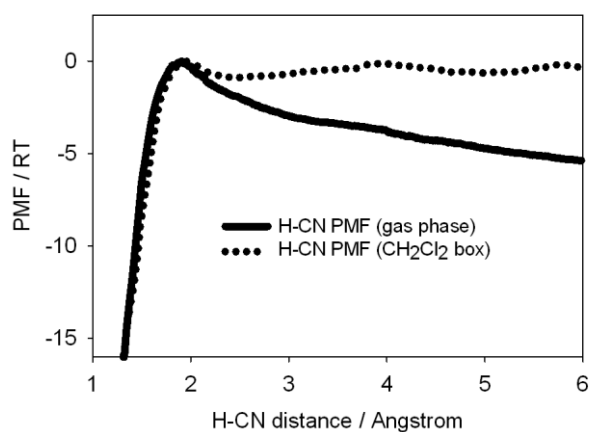
For these studies, the utility of BXD arises from the fact that it provides an integrated methodology for (1) efficiently mapping the solution phase PMF, and (2) accelerating reactive events that would otherwise not be possible to simulate. To get PMFs for these solution phase systems, ‘boxes’ were defined in terms of the distance between the approaching CN and an H atom on the co-reactant. Figure 7c shows results for CN + C₆H₁₂.

PMF sampling in BXD is particularly elegant in these sorts of systems; however, the principal advantage of BXD for the aforementioned solution phase systems is its ability to provide rare event acceleration that furnishes meaningful dynamics insight. These studies were carried out by placing a BXD box boundary to the reactant side of the transition state, and ‘locking’ the trajectory so as not to go beyond this boundary toward reactants. This approach⁹ guarantees that the system does not stray too far toward isolated reactants, and simultaneously accelerates transition state passage. So long as the BXD box boundary is far enough away from the transition state, then the post-transition state dynamics are decorrelated from dynamical inversion events and provide statistically meaningful dynamics results.^{2, 9} Conveniently, this allows us dynamical insight into how the system behaves unperturbed by BXD box.



(a)

(b)



(c)

Figure 6 Panel (a) shows a schematic of CN + cyclohexane in dichloromethane solvent; panel (b) shows CN + THF in THF solvent; panel (c) shows the PMF calculated for CN + cyclohexane

The high-quality rare event statistics available with the BXD technique have permitted several new insights into solution phase dynamics. One of the most interesting is the direct observation of two distinct post-reaction non-equilibrium relaxation regimes^{19, 39}. For example, in the case of the CN + C₆H₁₂ reaction in CH₂Cl₂ solvent, the first relaxation regime follows in the immediate wake of the reaction, when the nascent HCN is in close proximity to

its cyclohexyl coproduct, and strong interaction between co-products facilitates rapid energy transfer and correspondingly fast HCN relaxation. The second relaxation regime occurs once HCN and cyclohexyl have diffused away from one another and into the bulk solvent. HCN relaxation is considerably slower following diffusion, owing to weaker interactions between HCN and the CH_2Cl_2 solvent.

For the systems shown in Fig. 6, the potential energy surfaces for reaction were obtained using a multiconfigurational molecular mechanics approach, with Hamiltonian matrix elements fitted to configuration space points along the reaction path, calculated using high level CCSD(T) electronic structure theory results extrapolated to the infinite basis set limit. This approach provides a means for smoothly and accurately connecting molecular mechanics forcefields that describe both the reactant and product configuration. QM/MM approaches would have similarly allowed us to treat the reactive process, but at a considerably increased computational cost.

2.5 Free energy of unfolding

In a number of recent experiments³²⁻³⁴ the unfolding of single protein molecules with mechanical force has been investigated using the Atomic Force Microscope (AFM). In one class of these experiments, the proteins are pulled at a constant speed, (anywhere between 40 and 4000 nm s⁻¹); in another class, the proteins are pulled at a constant force. The net result of such experiments is a large quantity of experimental data with detailed information on force as a function of peptide extension.

Since the timescale of the pulling experiment is quite long, atomistic MD simulations tend to utilize pulling speeds and forces which exceed those of experiment by many orders of

magnitude. BXD may be easily applied to simulations of AFM owing to the fact that the unfolding process is reasonably well described using a simple end-to-end distance order parameter (similar to the peptide cyclization example considered above). This allows us to investigate events which would otherwise be far too expensive to simulate. In what follows, we present the first preliminary results of BXD simulations applied to understand AFM unfolding experiments. We calculated free energies along the reaction coordinate for 3 proteins, all of which are shown in Figure 7: IM9 is an all-alpha protein; I27 is an all-beta protein; and Protein L is an alpha-beta protein. Experimentally, the AFM results showed a correlation between the mechanical strength of the proteins and their secondary structure, with β -proteins requiring a much higher unfolding force than α -helical proteins.

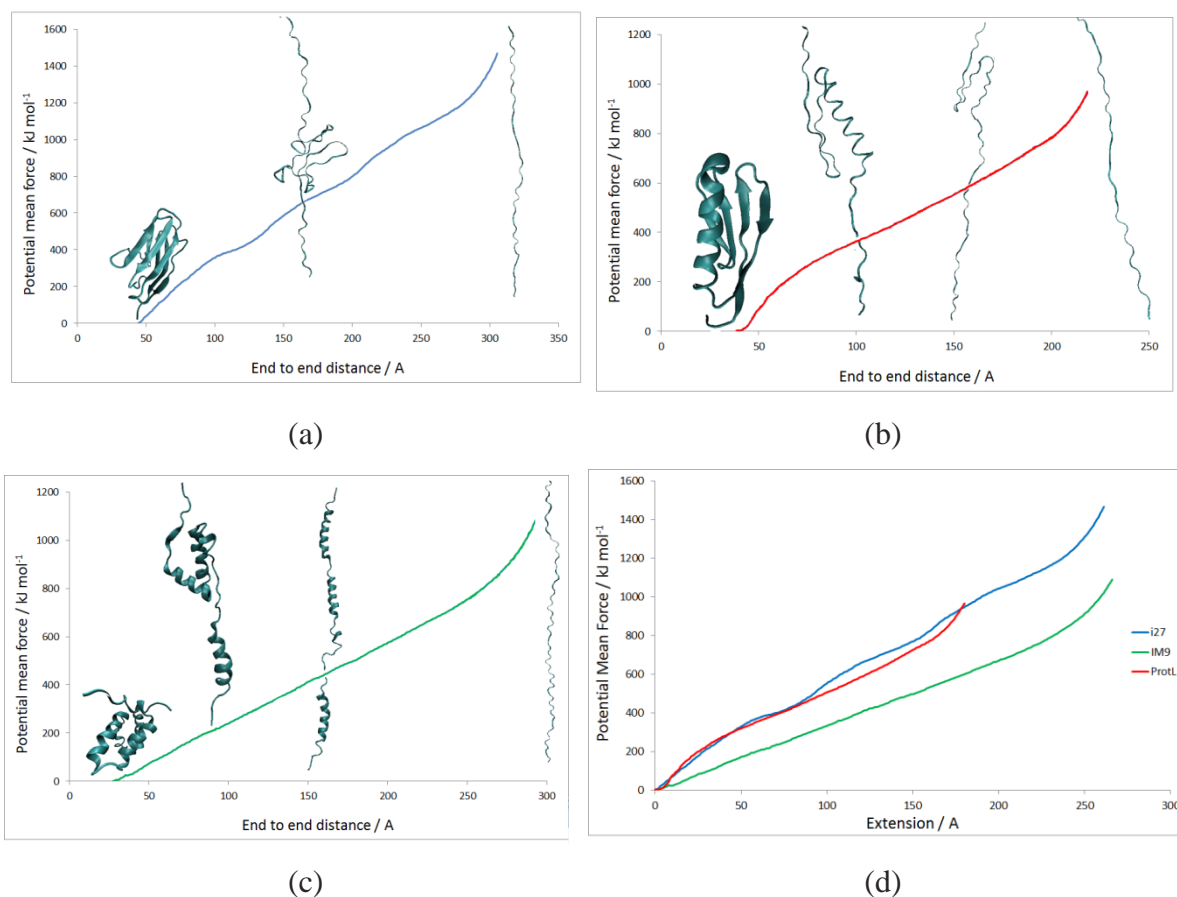


Figure 7 Potential of Mean Force profile as a function of end-to-end distance for the proteins (a) I27, (b) IM9, and (c) L. Panel d shows a comparison of the respective PMFs. The slope of free energy is significantly smaller for protein IM9, which has only α helical secondary structure. Snapshots of structures along the PMF profile are also shown.

Figure 7 shows PMF as function of end to end distance with results obtained using the previously discussed decorrelation algorithm.² The extension coordinate of proteins IM9, I27 and L were split into uniform boxes with sizes of 3 Å, 2 Å and 1 Å, respectively. The average derivatives $\frac{dG}{dx}$ (calculated by finite difference over an interval of 20 Å) in Figure 7 are as follows: ~65 pN for protein Im9, ~170 pN for ProteinL, and ~170 pN for I27. These values are in broad agreement with trends reported for the experimental unfolding forces: ~20 pN for Im9, 185-200 pN for I27 and 100-200 pN for ProteinL (depending on the pulling speed).³²⁻³⁴

These results are preliminary, but nevertheless show qualitative agreement with the experimental trend. In future work we will perform an extensive range of checks in an attempt to unravel the unfolding pathways, and determine how the dynamical pathways provided by BXD map into simulations compare to simulations carried out with an applied force. We will also perform kinetic studies of the extension dynamics and obtain time dependent forces and distances under various experimental conditions – e.g., varying the pulling speed and the direction of pulling, which should permit a more direct comparison with experiment. For example, if a protein is pulled by the force F then it is possible to calculate a force-modified PMF by adding the term $-Fx$ to the PMF, and correcting the corresponding the box-to box rate constants by the factor $\exp\left(\pm \frac{F\Delta x}{2kT}\right)$ where Δx is the box size. Initial kinetic simulations indicate that end-to-end distance starts increasing and a protein begins unfolding when the force becomes comparable to the slope of the free energy plot.

3. Conclusions and future developments

In this article we present a brief summary of the Boxed Molecular Dynamics (BXD) method, which allows the simultaneous calculation of thermodynamics (free energies along reaction coordinate) and kinetics of complicated molecular systems. We review previous applications of BXD across a range of different chemical systems, such as peptide cyclization and solution phase bimolecular reaction dynamics. We also report first results obtained from simulating desorption of ions from the surface of a self-assembled monolayer, the kinetics of diamond etching, and preliminary results of simulations of protein unfolding in AFM. Atomistic simulations of these processes by standard MD are not possible, but BXD yields accurate and physically meaningful results. BXD is a simple method which is easy to implement, and which can accurately calculate free energies along a reaction coordinate or some appropriately partitioned region of molecular phase space. BXD allows one to recast the complicated atomistic dynamic evolution a given system as a set of kinetic processes occurring within different regions of configuration space. Although BXD is a fully atomistic technique, it is capable of describing dynamics many orders of magnitude longer than typical atomistic dynamics timescales. The systems outlined in this article demonstrate how BXD has furnished considerable microscopic insight into the atomistic processes and mechanisms at work in a range of condensed phase and heterogeneous systems.

A number of BXD projects are currently underway in our groups, including the dynamics of diamond etching, additional solution phase reaction systems, the dynamics of FM unfolding, the dynamics of pocket opening in biomedically important proteins (where access to the docking site is blocked by structural elements of the protein), and the folding dynamics of a range of small proteins. In a number of these systems, we obtain free energy/PMF profiles as

well as the long-time kinetics / dynamics, furnishing considerable microscopic insight into a range of problems. The applications reported and reviewed in this paper have utilized relatively simple reaction coordinates, but we are presently exploring more sophisticated reaction coordinates, as well as multidimensional extensions of BXD.

Acknowledgements

DRG is grateful for funding from a Royal Society Research Fellowship. JB and DVS acknowledge the support of EPSRC (Grant No EP/E009824/1). E.M.-N. and S.A.V. are grateful to the “Centro de Supercomputación de Galicia (CESGA)” for the use of its computational resources, as well as to “Ministerio de Economía y Competitividad” (Grant No. CTQ2009-12588) for financial support. DS and E.M.-N. acknowledge the Leverhulme Trust for funding the E.M.-N. visit to Leeds by the grant “Accelerated classical and quantum molecular dynamics and its applications” (Grant No. VP1-2012-013).

References

- 1 D. R. Glowacki, E. Paci and D. V. Shalashilin, *Journal of Physical Chemistry B*, 2009, **113**, 16603-16611.
- 2 D. R. Glowacki, E. Paci and D. V. Shalashilin, *Journal of Chemical Theory and Computation*, 2011, **7**, 1244-1252.
- 3 D. V. Shalashilin, G. S. Beddard, E. Paci and D. R. Glowacki, *Journal of Chemical Physics*, 2012, **137**, 9.
- 4 Y. Guo, D. V. Shalashilin, J. A. Krouse and D. L. Thompson, *J. Chem. Phys.*, 1999, **110**, 5521-5525.
- 5 Y. Guo, D. V. Shalashilin, J. A. Krouse and D. L. Thompson, *J. Chem. Phys.*, 1999, **110**, 5514-5520.
- 6 D. V. Shalashilin and D. L. Thompson, *J. Phys. Chem. A* 1997, **101**, 961-966.
- 7 D. V. Shalashilin and D. L. Thompson, *J. Chem. Phys.* , 1997, **107**, 6204-6212.
- 8 D. V. Shalashilin and D. L. Thompson, *ACS Symp. Ser.* , 1997, **678**, 81-98.
- 9 E. Martinez-Nunez and D. V. Shalashilin, *J. Chem. Theory Comput.* , 2006, **2**, 912-919.
- 10 R. Elber, *Biophysical Journal*, 2007, **92**, L85-L87.
- 11 A. K. Faradjian and R. Elber, *J. Chem. Phys.* , 2004, **120**, 10880-10889.
- 12 K. Kuczera, G. S. Jas and R. Elber, *J. Phys. Chem. A*, 2009, **113**, 7461-7473.
- 13 L. Maragliano, E. Vanden-Eijnden and B. Roux, *Journal of Chemical Theory and Computation*, 2009, **5**, 2589-2594.
- 14 E. Vanden-Eijnden, M. Venturoli, G. Ciccotti and R. Elber, *J. Chem. Phys.*, 2008, **129**, 174102/174101-174102/174113.
- 15 A. M. A. West, R. Elber and D. Shalloway, *J Chem Phys*, 2007, **126**, 145104.
- 16 E. Vanden-Eijnden and M. Venturoli, *J. Chem. Phys.* , 2009, **130**, 194101/194101-194101/194113.
- 17 D. R. Glowacki, R. A. Rose, S. J. Greaves, A. J. Orr-Ewing and J. N. Harvey, *Nature Chemistry*, 2011, **3**, 850-855.
- 18 S. J. Greaves, R. A. Rose, T. A. A. Oliver, D. R. Glowacki, M. N. R. Ashfold, J. N. Harvey, I. P. Clark, G. M. Greetham, A. W. Parker, M. Towrie and A. J. Orr-Ewing, *Science*, 2011, **331**, 1423-1426.
- 19 D. R. Glowacki, A. J. Orr-Ewing and J. N. Harvey, *Journal of Chemical Physics*, 2011, **134**, 11.
- 20 A. F. Voter, *Physical Review Letters*, 1997, **78**, 3908-3911.
- 21 A. F. Voter, *Journal of Chemical Physics*, 1997, **106**, 4665-4677.
- 22 G. M. Torrie and J. P. Valleau, *Journal of Chemical Physics*, 1977, **66**, 1402-1408.
- 23 G. M. Torrie and J. P. Valleau, *Journal of Computational Physics*, 1977, **23**, 187-199.
- 24 J. Kaestner, *Wiley Interdisciplinary Reviews-Computational Molecular Science*, **1**, 932-942.
- 25 D. Frenkel and B. Smit, *Understanding Molecular Simulations*, , 2002, **2nd ed**, Academic Press, London.
- 26 R. Metzler, J. Klafter, J. Jortner and M. Volk, *Chemical Physics Letters*, 1998, **293**, 477-484.
- 27 M. Volk, Y. Kholodenko, H. S. M. Lu, E. A. Gooding, W. F. DeGrado and R. M. Hochstrasser, *Journal of Physical Chemistry B*, 1997, **101**, 8607-8616.
- 28 L. Milanese, J. P. Waltho, C. A. Hunter, D. J. Shaw, G. S. Beddard, G. D. Reid, S. Dev and M. Volk, *Proceedings of the National Academy of Sciences of the United States of America*, 2012, **109**, 19563-19568.
- 29 H. Luo, S. A. Miller, R. G. Cooks and S. J. Pachuta, *International Journal of Mass Spectrometry*, 1998, **174**, 193-217.
- 30 S. A. Miller, H. Luo, X. Jiang, H. W. Rohrs and R. G. Cooks, *International Journal of Mass Spectrometry and Ion Processes*, 1997, **160**, 83-105.
- 31 S. A. Miller, H. Luo, S. J. Pachuta and R. G. Cooks, *Science*, 1997, **275**, 1447-1450.

- 32 D. J. Brockwell, *Biochemical Society Transactions*, 2007, **35**, 1564-1568.
- 33 E. Hann, N. Kirkpatrick, C. Kleanthous, D. A. Smith, S. E. Radford and D. J. Brockwell, *Biophysical Journal*, 2007, **92**, L79-L81.
- 34 D. P. Sadler, E. Petrik, Y. Taniguchi, J. R. Pullen, M. Kawakami, S. E. Radford and D. J. Brockwell, *Journal of Molecular Biology*, 2009, **393**, 237-248.
- 35 B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545-1614.
- 36 J. J. Nogueira, Y. Wang, M. F., M. Alcamí, D. R. Glowacki, D. V. Shalashilin, E. Paci, W. Hase, L., E. Martinez-Nunez and S. Vazquez, *Submitted*.
- 37 D. R. Glowacki, W. J. Rodgers and J. N. Harvey, *in preparation*.
- 38 A. J. Orr-Ewing, D. R. Glowacki, S. J. Greaves and R. A. Rose, *Journal of Physical Chemistry Letters*, 2011, **2**, 1139-1144.
- 39 D. R. Glowacki, R. Lightfoot and J. N. Harvey, *Molecular Physics*, 2013, **111**, 631-640.
- 40 R. A. Rose, S. J. Greaves, F. Abou-Chahine, D. R. Glowacki, T. A. A. Oliver, M. N. R. Ashfold, I. P. Clark, G. M. Greetham, M. Towrie and A. J. Orr-Ewing, *Physical Chemistry Chemical Physics*, 2012, **14**, 10424-10437.